Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 50 (2015) 635 – 642

Procedia Computer Science

2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Conceptual Graph Representation Framework for ICD-10

N. Hema[a], S. Justus[b]

[a,b] SCSE, VIT University, Chennai

Abstract

Codification and representation of knowledge from structured and unstructured repositories have been the requirement for Expert Systems designing. Structured repositories remain open for knowledge extraction and representation for building a knowledgebase (KB). ICD-v10 (International Classification of Diseases), a structured repository, is the 10th revision of the collection of Disease codes and health related problems, symptoms, complaints and external causes of injury. ICD-v10 uses a 7-bit code to represent all diseases. The main focus of our research work is to propose new methods for capturing and representing these codes in the forms of Knowledge from Semantic Structured Repository with better materialization, quality and repository usage, along with logical unit and logical entailments to form conceptual relations among them. Apart from creating a new representation structure, this work gives direction on the storage of these structures as Knowledge Unit (KU).

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Keywords: Knowledge Base (KB); OWL; Conceptual Graph; Description Logic, ICD-10; First Order Logic (FOL); Knowledge Representation

1. Introduction

Medical coding is the process of converting medical diagnosis and procedures into a unique universal medical code number. These derived medical codes are used by the medical insurance company for reimbursement process, for statistical analysis of diseases, the actions and procedure adopted in the US and European countries. The obtained statistical results help in identifying the people ratio affected by a particular symptom at a particular location and season. This may help the Health Ministry of a country to eradicate or control any kind of

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)
doi:10.1016/j.procs.2015.04.097

communicable diseases in the location.

The medical codes are classified based on [18]

- Diagnosis codes ICD-10 CM (International Classification of Disease-10 Clinical Modification)
- Procedural codes ICD-10 PM (International Classification of Disease-10 Procedural Modification), ICHI (International Classification of Health Intervention), CCHI (Canadian Classification of Health Intervention), ACHI (Australian Classification of Health Intervention), HCPCS (Health Care Procedure Coding System)
- Pharmaceutical codes AT (Anatomical Therapeutic), NDC (National Drug Code)
- Topographical codes ICD-O (International Classification of Disease – O), SNOMED (Systematized Nomenclature of Medicine)

1.1. Prior to ICD-10

The Family of International Classification (FIC) has three main reference based on the International health parameter, along with some relative classification with additional information. There are some reference classifications for international disease;

Table 1. Medical Codes Classification such as ICD

CDT – Code of Dental Procedures
CPT – Current Procedural Terminology
DRG – Diagnosis Related Groups
DSM – Diagnosis and Statistical Manual of Mental Disorders
HCPCS – Healthcare Common Procedure Coding System
HCUP – Healthcare Cost and Utilization Project
HL7 – Health Level Seven
ICF – International Classification of Functioning, Disability and Health
LOINC – Logical Observation Identifier Names and Codes Terminology
UMLS – Unified Medical Language System, NLM
SNOMED CT – Systematized Nomenclature of Medicine – Clinical Terms

1.2. What is ICD?

The ICD-10-CM/PCS is an International Unique Classification of Disease followed by medical practitioners. The diseases in ICD are classified based on various factors such as severity and the degree or level to which a disease can affect the human body. There are many versions of ICD, and the current version is ICD-10 CM/PCS. This version has almost 68,000 clinical diagnosis area codes and 87,000 codes for operations and procedures, providing plenty of room for expansion [11]. The ICD codes can have three, four, five, six or seven characters in their evolutionary ICD versions [19]. The three character codes act as category headings and can further be expanded up to seven characters to provide more additional details of diagnosis.

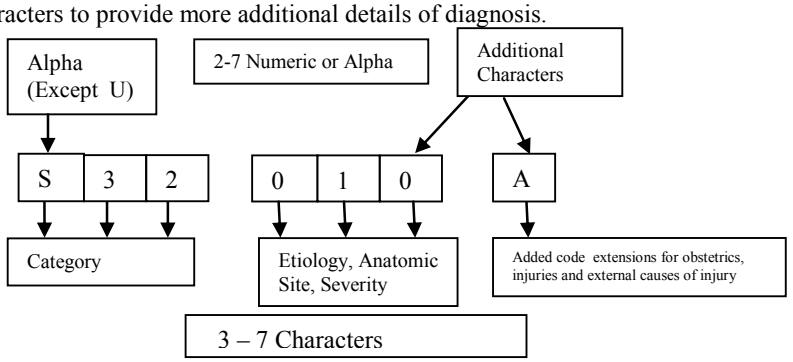

Fig.1 ICD-10-CM Structure

The first three characters decide the category of the diagnosis. The first digit, the character 'S' stands for "Injuries, Poisoning and certain other consequences of external causes related to single body regions". The second and third digits, integer values '3' and '2' in conjunction with 'S' means "Fracture of lumbar spine and pelvis"[11]. The conjunction of fourth and fifth character S32.01 means "Fracture of first lumbar vertebra". The sixth character specifies the severity and the extended seventh character describes whether it be an open or closed fracture.

Eg: The code *H40.11X1* describes "*Mild stage primary open-angle glaucoma*"

Similar to the ICD-10-CM version, we also have a separate structural format for ICD-10-PCS version also. Each of the 7 digits can have any 34 values (0 to 9, A – Z except O & I).

H40.11 Primary open-angle glaucoma
Chronic simple glaucoma

One of the following 7th characters is to be assigned to code H40.11 to designate the stage of glaucoma
0 - stage unspecified
1 - mild stage
2 - moderate stage
3 - severe stage

Fig 2. (a) ICD-10 Code Classification

1	2	3	4	5	6	7
Section	Body System	Root Operation	Body Part	Approach	Device	Qualifier

Fig 2.(b) ICD-10-PCS Structure

The first three digits in fig. 8 (b) describe the section, body system and root operation of the procedure of diagnosis. The fourth position is specific about the atomic part where the procedure was done. The fifth position details the technique used for the procedure (Open/ Closed/ Laproscopy/ ..). The sixth position specifies the device used for the technique and the seventh position gives the additional attribute of the procedure [12].

2. Related Work

ICD-10-CM is available from the CDC FTP server, in a structured XML data format, as well as PDF for viewing [20]. Also, for a fixed width text format that includes primarily just the codes and titles from the tabular part of ICD-10-CM, the description order files could be useful. There is not a database version, but the fixed width description order files could be readily loaded into a database system.

A variant representation of the tabular part of ICD-10-CM has been made available by the National Library of Medicine, and is available in the UMLS, [21] which also includes additional information in a format that can more readily be loaded into database systems and other applications. However, it requires significant computer resources to install the UMLS and to extract this information. It also requires an agreement for access, because it also includes other proprietary terminologies.

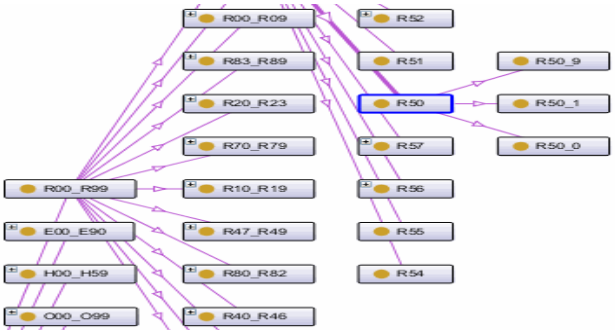

Fig 3. Tree Structural representation of ICD-10 CM data in Protégé

The structured XML data format of the ICD-10-CM can be converted to an OWL structure using Protégé. This gives tree structural representation of the ICD-10-CM data helping to understand just the classification of the main term to sub terms alone with the diagnosis. Even though the OWL Ontology Graph shown in Fig 3, provides us with sufficient information like the sub classes, super classes, disjoint classes and annotation of the diagnosis as given in [5] there are some disadvantages.

The existing XML format does not provide sufficient information about the Include and Exclude codes and the Additional codes for higher level of specificity. Also the database loading facility provided by the Centre of Disease Control (CDC) File Transfer Protocol server, maintains a fixed width description of the codes. This may lead to incompatibility in future, when there may occur a need for extra characters to be added to the ICD code, as in ICD-10 transition. Moreover, in the manual system of diagnosis of code, identification of main term, sub term or more specific term is not easy, as it is based on the way the human knowledge interprets. More chances are there to run with the risk of wrong diagnosis of the ICD codes.

2.1. Basic Coding Steps

Coding of medical diagnosis Code is a very difficult and a challenging task.

- First identify the main term in the diagnosis statement.
- Locate the main term in the Alphabetic Index and review any sub-term or more specific sub-term.
- Verify the code number in the Tabular List [15].
- Assign the verified code or codes to the highest level of specificity.

Table 2. Main categories of the Classification of ICD-10 Codes

A00-B99	Certain infectious and parasitic diseases
C00-D49	Neoplasms
D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E89	Endocrine, nutritional and metabolic diseases
F01-F99	Mental, Behavioral and Neuro developmental disorders
G00-G99	Diseases of the nervous system
H00-H59	Diseases of the eye and adnexa
H60-H95	Diseases of the ear and mastoid process
I00-I99	Diseases of the circulatory system
J00-J99	Diseases of the respiratory system
K00-K95	Diseases of the digestive system
L00-L99	Diseases of the skin and subcutaneous tissue
M00-M99	Diseases of the musculoskeletal system and connective tissue
N00-N99	Diseases of the genitourinary system
O00-O9A	Pregnancy, childbirth and the puerperium
P00-P96	Certain conditions originating in the perinatal period
Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
S00-T88	Injury, poisoning and certain other consequences of external causes
V00-Y99	External causes of morbidity
Z00-Z99	Factors influencing health status and contact with health services

For the following diagnosis, the coding is done as;

Fever - Since fever is the main term, we locate the term Fever in the Alphabetic Index. [14]

The code number for the term fever is R50.9. We then verify this code with the Tabular list. Both the code in the Alphabetic and Tabular index matches and hence R50.9 is assigned for the diagnosis fever. There may be one

or more ICD code for a single diagnosis. This way of manual solution to the problems are done by both common sense and specialized knowledge that gives the meaning of the diagonized data. Hence there are more chances of performance limitation or quality of the result.

Fever (inanition) (of unknown origin) (persistent) (with chills) (with rigor) R50.9
- abortus A23.1
- Aden (dengue) A90
- African tick-borne A68.1
- American
- - mountain (tick) A93.2
- - spotted A77.0

Fig 4. (a) Alphabetic Index Representaion for Fever

Source : http://www.cdc.gov/nchs/icd/icd10.htm

R50 Fever of other and unknown origin
R50.2 Drug induced fever
R50.8 Other specified fever
R50.81 Fever presenting with conditions classified elsewhere
R50.82 Post procedural fever
R50.83 Post vaccination fever

Fig 4. (b) Tabular list representation for Fever

3 Proposed System

The main objective of the proposed system is to provide a new knowledge representation and coding support tool for the International Classification of Diseases (ICD-10). In this method we propose a step by step process for storing the ICD code and its actual relationship into the database, rather than directly loading from the server. This helps in identifying the diagnosis code of any given symptoms easily and accurately [4].

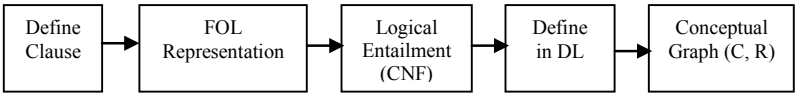

Fig 5. Proposed Knowledge Representation of ICD-10

First, we construct clauses from the observed symptoms. These clauses are represented using the FOL format using the conditions and description of the ICD-10 CM. They are further normalized using CNF for Knowledge entailment and reasoning. This is then converted into a Descriptive Logic by adding logical inferences of the past experience and real world knowledge. Finally, to effective manipulation and decision making we will use Conceptual Graph (CG) way of reasoning and logical conclusions.

3.1 Defining a Clause

A clause is a group of word having a verb and other components. There are different kinds of clause a verb may be termed. The identified verb may be a;

- Main Clause
- Subordinate Clause
- Relative Clause
- Restrictive Relative Clause
- Non-restrictive Clause
- Conditional Clause

i) Main Clause

There exists at least one main clause in a sentence. This main clause may be a part of complex sentence. Complex sentences have two or more main clause connected using and, or, but, with, so etc.

Examples of Compound statements:

- Diabetes with Ketoacidosis (where diabetes and ketoacidosis are two main terms)
- Obesity and other hyperalimentation (where obesity and Hyperalimentation are two main term)

- Typhoid and paratyphoid fevers (where typhoid and fever are two main terms)
- Chronic Kidney diabetes (where diabetes and chronic kidney are the two main terms)

ii) Subordinate Clause

The subordinate clause depends on the main clause to form a complete meaningful statement. These subordinate clauses cannot be the manifestation code for any underlying disease and are represented using the term ***"Code First"*** in the ICD-10 representation format.

Example:

Statement	**Main Clause**	**Subordinate Clause**
Tension headache	Headache	Tension
Disorders of Thyroid gland	Disorder	Thyroid Gland
Tuberculosis of Nervous System	Tuberculosis	Nervous System

iii) Relative Clause:

Relative clause is always connected to the main clause using special keywords like when, where, due to, that …

Statement	**Main Clause**	**Relative Clause**
Alzheimer due to Dementia	Alzheimer	Dementia
Infection due to other mycobacteria	Infection	Mycobacteria
Stage IV type 1 diabetes due to chronic kidney disease	chronic kidney disease	Stage IV type 1 diabetes
Arthritis due to Lyme disease	Lyme disease	Arthritis
Syphilis of bone and joint	bone and joint	Syphilis

These relative clauses can either be restrictive or non-restrictive. The restrictive relative clause gives additional information, without which the sentence is meaningless. These codes are represented using the term ***"Use additional code"*** or ***slanted Brackets [ ]*** in Alphabetic Index of ICD coding format. A non-restrictive relative clause also provides additional information to the noun, but does not affect the meaning of the sentence. These type of codes are written inside ***parenthesis ()*** in ICD code.

Eg: J18.9-Pneumonia (acute)(double)(migratory)(septic) (purulent)

iv) Conditional Clause:

A conditional clause describes something that is possible or probable due to some condition. They may be a single condition or a group of condition for a possibility to occur.

3.2 Defining a First Order Logic

FOL is a format mathematical system that symbolizes reasoning where each sentence of statement is categorized into a subject and a predicate. The property of the subject is modified or defined by the predicate. These predicates refer to one subject. FOL are more powerful forms of logic and acts as a good reasoned.

Eg:
a) $\forall x($ diabetes$(x) \Lambda$ ketoacidosis$(x) \Lambda \neg$coma$(x)) \Rightarrow$ E10.10
b) $\forall x($ diabetes$(x) \Lambda$ ketoacidosis$(x) \Lambda$ coma$(x)) \Rightarrow$ E10.11
c) $\forall x \exists y($Pneumonia$(x) \Lambda$ acute$(y) V$ double$(y) V$ migratory$(y) V$ septic$(y) V$ purulent$(y) V$ unresolved$(y) V \neg y) \Rightarrow$ J18.9
d) behaviour(y)
$\forall x \exists y ($ dementia$(x) \Lambda \neg ($ aggressive$(y) V$ disturbance$(y) V$ combative$(y)) \Rightarrow$ F03.90
$\forall x \exists y ($ dementia$(x) \Lambda ($ aggressive$(y) V$ disturbance$(y) V$ combative$(y)) \Rightarrow$ F03.91
e) $\forall x ($ Diabetes$(x) \Lambda$ type-1$(x) \Lambda (\exists y$ disease$(y) \Lambda$ chronic-kidney$(y))) \Rightarrow$ E10.22

3.3 Defining the Conjunctive Normal Form

The advantage of converting FOL to Conjunctive Normal Form (CNF) is to provide Logical Entailment and make sure that certain queries can be answered easily and accurately[16]. This is usually written as "Product of Sum". Here, the rule is said to be Satisfiable, Logically Valid or Unsatisfiable. A CNF formula is said to be a

tautology, if the sum has a variable and a negation.

3.4 Defining a Description Logic

FOL is not decidable and finding a proof for these statements takes more time. Whereas, a Description Logic is a decidable fragment, with an efficient reasoned RACER[2]. It is used for formal reasoning based on the domain in Artificial Intelligence. It has various forms on an expression as;

- Concepts like nouns
- Roles like relational nouns
- Constants like proper noun

3.5 Defining a Conceptual Graph:

Conceptual Graph is a form of knowledge representation to represent the concept structure in a database. Concepts are the combination of words; and the way it is linked together with other concepts forms knowledge. The concepts are classified into two kinds of expressions namely, types (classes) and values[1]. V(G) = <c, r >

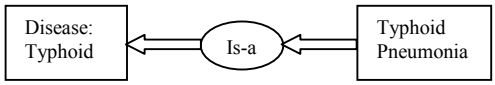

Fig. 6 Graph based representation with is-a relation

Here, Typhoid is the value and type is the Disease. CG allows a type-value pair concept. “Typhoid Pneumonia” is the possible value of the of the left hand side type. Similarly “Typhoid Arthritis” can also become the right hand value. They have two different kinds of nodes in its representation namely; concepts and relations[10]. The relations are the role of the concepts having the actual knowledge. There are two kinds of basic relationship in CG. They are;

- The hierarchical relation (= "is-a") is represented horizontally. [1]
- The attributive relation (= "has-a") is represented vertically.[1]

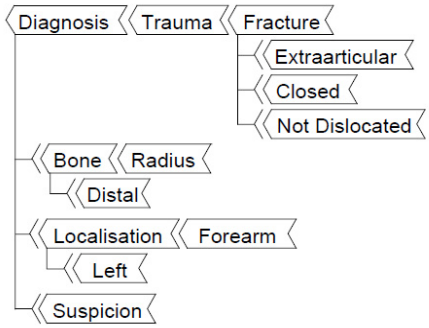

Fig 7. Hierarchical Graph representation of Fracture

Conceptual Graph’s can be represented in three different ways, namely; Deterministic Automatic Transformation, Statistical Automatic Transformation and Manual Transformation.[4]

4 Conclusions and Future Work

We have shown how good the new form of Conceptual Graph representation would visually show the requirements of knowledge, than a normal OWL representation in term of logical entailment and representation into a database. Also, the usage of Description Logic Reasoner [7] in-terms of concepts and roles. Also, we have tried to represent a new representation model for transforming a given simple sentence to a knowledge and store it in a database.

In future, we plan to work on an interface through which new updates and modifications to the Graph representation of knowledge could be made. We will also try to experiment with other alternate similar terms represented by physicians, and match it to the existing system. Through this, we are trying to bring in a hand held application for the physicians, to have a quick reference. This may also help us to make geographical and climatic analysis of any disease and their cause and remedies.

References

1. J. F. Sowa, "Conceptual Structures-- Information processing in mind and machine", Addison-Wesley Systems Programming Series Reading, MA, 1984
2. F. Bader et al. (editors): The Description Logic Handbook (Theory, Implementation and Applications), Cambridge University Press, 2003
3. Anunchai A., Thepchai S, Sissades. A Development of Knowledge Representation for Thalassemia Prevention and Control Program. T. 978-1-61284-729-0/11. 2011 IEEE.
4. Sonia Ordoñez-Salinas, Alexander; Gelbukh, " Information Retrieval with a Simplified Conceptual Graph-Like Representation"
5. Gergely Héja, György Surján, Gergely Lukácsy, Péter Pallinger, Miklós Gergely. GALEN based formal representation of ICD 10, International Journal of Medical informatics, Volume 76, Issues 2-3, Pages 118–123
6. Stefan F., Simon P, A Community Knowledge Base for IT Security. 1520-9202/11/IEEE Computer Society.
7. Ronald J. Brachman, Hector Levesque. Knowledge Representation and Reasoning. Edition 2010.
8. Bernard Kamsu-Foguem , Gayo Diallo , Clovis Foguem. Conceptual graph-based knowledge representation for supporting reasoning in African traditional medicine, http://dx.doi.org/10.1016/j.engappai.2012.12.004
9. B. R. Duffy, G. M. P. O'Hare, R. P. S. O'Donoghue, C. F. B. Rooney, and R. W. Collier, "Reality and Virtual Reality in Mobile Robotics, PRISM Laboratory," Dept. Comput. Sci., Univ. College Dublin, Dublin, Ireland, 2013. [Online]. Available: http://chameleon.ucd.ie /publications/manse99.ps
10. Hans Rudolf Strauba, Norbert Freib, Hugo Mosimanna, Csaba Pergera, Annette Ulricha aSemfinder AG, Kreuzlingen, Schweiz bUniversity of Applied Sciences St. Gallen (FHS), St. Gallen Schweiz. Simplified Representation of Concepts and Relations on Screen
11. ICD-10 Volume 2 Instruction Manual 2010 Edition.
12. Richard F. Averill, M.S., Robert L. Mullin, M.D., Barbara A. Steinbeck, RHIT, Norbert I. Goldfield, M.D, Thelma M. Grant, RHIA, Rhonda R. Butler, CCS, CCS-P. Development of the ICD-10 Procedure Coding System (ICD-10-PCS)
13. http://www.icd10data.com/ICD10CM/Codes. Accessed 12 July 2014
14. http://www.icd10forkindergarten.com/download. Accessed 15 August 2014
15. https://www.cms.gov/Medicare/Coding/ICD10/.../ ICD10Introduction.pdf
16. http://en.wikipedia.org/wiki/Conjunctive_normal_form. Accessed 7 December 2014
17. http://www.springer.com/978-0-85729-298-8
18. World Health Organization. Family of International Classifications. Accessed 12 July 2014
19. Brooke Andrus. Understanding the ICD -10 Code Structure.
20. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD10CM/2015/. Accessed 5 August 2014
21. http://www.nlm.nih.gov/research/umls/. Accessed 5 August 2014